Abstract Title Page
Not included in page count.

Title: Using Generalized Additive Models to Analyze Single-Case Designs

Authors and Affiliations:

William Shadish
University of California, Merced
School of Social Sciences, Humanities and Arts
5200 N. Lake Rd
Merced CA 95343
209-228-4372
wshadish@ucmerced.edu

Kristynn Sullivan
University of California, Merced
School of Social Sciences, Humanities and Arts
5200 N. Lake Rd
Merced CA 95343
ksullivan7@ucmerced.edu

Acknowledgment:
This research was supported in part by grants R305D100046 and R305D100033 from the Institute for Educational Sciences, U.S. Department of Education, and by a grant from the University of California Office of the President to the University of California Educational Evaluation Consortium.

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context.

Many analyses for single-case designs (SCDs)—including nearly all the effect size indicators—currently assume no trend in the data. Regression and multilevel models allow for trend, but usually test only linear trend and have no principled way of knowing if higher order trends should be represented in the model. This paper shows how Generalized Additive Models (GAMs) can be used to inform this aspect of the analysis.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

It is well known in traditional interrupted time series analysis that correctly modeling trend in the data is essential to obtaining an accurate effect size estimate. Nonlinearities can be inherent in the data (e.g., weight loss that slows over time resulting in a quadratic trend) or can result from interactions of the treatment with time (e.g., after a stable baseline, a treatment slowly becomes more effective as more sessions occur). Modeling trend is difficult in SCDs because they rarely have the large number of observations on a case over time to allow use of traditional methods such as ARIMA modeling. Analyses of SCDs by ordinary regression or multilevel models can address trend, but require the researcher to impose the particular functional form, something they can only intuit from visual inspection of the graph. GAMs are a semi-parametric regression model that allows the data to inform the required functional form. We show how they can be applied to SCDs. Given the early stage of this research, we propose the use of GAMs as a sensitivity analysis for whether trend might affect the main conclusion about treatment effectiveness. We believe, however, they have potential to become a primary data analytic tool for SCD data.

Setting:

Description of the research location.
(May not be applicable for Methods submissions)

(Not applicable)

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.
(May not be applicable for Methods submissions)

(Not applicable)

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.
(May not be applicable for Methods submissions)

(Not applicable)

Significance / Novelty of study:
Description of what is missing in previous work and the contribution the study makes.

This work will inform SCD researchers about whether trends in their data are likely to be a problem for conclusions about treatment effects. This has obvious implications for the effect size measures that currently assume no trend. It also will greatly assist researchers using regression and multilevel models in deciding if higher order polynomial or interactions need to be included in their models—virtually all of them assume linear trend currently.

Statistical, Measurement, or Econometric Model:
Description of the proposed new methods or novel applications of existing methods.

GAMs are like generalized linear models (GLM) but they replace one or more of the usual terms of a GLM with a predictor that consists of a sum of smoothing functions (Wood, 2006). So GAM estimates both parametric and nonparametric terms. Terms with nonparametric smoothers (explained in more detail shortly) are used to test assumptions about trend: for instance, to see if the outcome is changing over time, whether change is linear, quadratic, or some other form, and whether the treatment effect remains constant over time as phases change. In principle, GAMs can be applied jointly to a set of several cases, incorporating random effects terms as in other hierarchical models, to assess the effects of an intervention on all cases in a study. However, we currently know too little about how to do this, and its strengths and weaknesses, to recommend GAM for that purpose in more than an exploratory sense. So here we focus on using GAMs to model each case separately. Our approach is consistent with the goal of using GAMs as a tool for model development and assumption-checking, rather than for summarizing results across an entire study. That approach will likely change as we learn more.

Within the GAM framework, the smoothing terms have to be represented in such a way that the GAM becomes a linear model. Imagine a simple case, in which you have one smoothed predictor: $Y_i = s(x_i) + \varepsilon_i$. Imagine further that you knew that s should result in a quadratic relationship. The equation for s would be: $s = \beta_1 + x\beta_2 + x^2\beta_3$. Substituting that in to the model equation, you get: $Y_i = \beta_1 + x_i\beta_2 + x_i^2\beta_3 + \varepsilon_i$, which is a linear model in the same fashion that all GLM predictors are linear, even in the presence higher order polynomial terms. The problem is that in the real world one doesn't know the true order of s. So, one chooses a basis, or a set of linearly independent vectors, that defines a functional space. These vectors, when linearly combined, can represent any potential vector in the basis space. All of the potential smoothing terms of the model are an element, or basis function, of the chosen basis. So, any potential smoothing term is some linear combination of linearly independent vectors in the basis. Choosing a basis allows the estimation of a nonlinear term from the data, but constrains the geometrical space from which they can be estimated (e.g. so that the smoothing does not result in an unrealistic value, such as a 100^{th} order polynomial smoother).

There are many potential basis options. A common basis is penalized cubic regression splines (CRS). Spline bases relate the smoothing function to the entire domain of data rather than a single point of the data. CRSs are constructed from pieces of cubic polynomial curves joined together into a continuous function. The curves are joined together at the knots of the data set; knots are the places where an inflection in the curve appears. CRSs are computationally efficient, and their results are easily interpretable. They also can be implemented on small data sets. With CRSs the researcher has to specify where to place knots, or the location of the potential bends in the functional relationship. One can choose to equally space these knots across the span of the

data. Generally, the computer program default spaces the knots evenly across the data, so this is not an arduous process.

Introducing smoothing parameters requires estimating the degree of smoothing necessary for each covariate, for example, the degree of smoothing necessary for the trend term in the present case. Each *s* term of a GAM model contains a smoothing parameter. The smoothing parameter estimates the optimal amount of smoothing to fit the data while simultaneously adding a penalty for increased "wiggliness" of the smoothing function. Adding a penalty matrix to the least squares estimation model avoids over-fitting the smooth to the data. Within this framework, *s* approaches a straight line as the smoothing parameter approaches infinity. The optimal degree of smoothness can be estimated directly from the data. GAMs are maximized by penalized iteratively re-weighted least squares (P-IRLS). The optimal smoothing parameter is chosen by calculating a generalized cross-validation (GCV) score of each iteration. The underlying idea is to remove one data point from the data set, re-estimate the model, and then estimate the predicted value of the removed data point, based on the new model. The observed data point is subtracted from the predicted value, and the deviations are squared. This process is repeated for every data point, and the squared deviations are averaged. The resulting average is the cross validation score. However, this process is computationally tedious. An equivalent score calculation is

$$V = n \sum_{i=1}^{n} (y_i - \hat{f}_i)^2 / [\mathrm{tr}(\mathbf{I} - \mathbf{A})]^2.$$

y_i is the observed data point, $\hat{f}$ is the predicted value of that data point, **I** is the identity matrix of the full model, and **A** is the model influence matrix, a matrix that maps the vector of observed values, **y**, to the vector of predicted or fitted values, **f**, and describes the influence that each observed value has on each fitted value. Tr(**I** - **A**) is the trace, or sum of the matrix diagonals, of the matrix that results from of the model influence matrix, **A**, subtracted from **I**, the identity matrix. The smaller the GCV score, the better the model fit. Models can be compared using their GVC scores, illustrated in the examples below.

The model output also lists the effective degrees of freedom of the smoothing term. The effective degrees of freedom is defined as the trace of **A**, the model influence matrix (recall that the influence matrix is a matrix that maps the vector of observed values to the vector of predicted or fitted values and describes the influence that each observed value has on each fitted value). Estimated degrees of freedom equal to one is a linear effect (Wood, 2006). As the effective degrees of freedom increase, the parameter smooth becomes wigglier. When using cubic regression splines, effective degrees of freedom are very roughly equivalent to the polynomial order of the smoother plus one. That is, effective degrees of freedom of 4 would roughly imply a third degree polynomial smoothing term for the covariate being tested (Hothorn & Everitt, 2010, Chapter 10). This makes interpretation of the nonlinearities of the predictors more intuitive, although the effective degrees of freedom are rarely whole numbers. This is also an extremely approximate rule of thumb, and as the effective degrees of freedom approach one, this rule of thumb no longer applies.

To simplify notation, we omit the subscript *i* used in previous sections to index each case, retaining only the index for time *t*. The basic model for a single-case is:

$$Log\left(\frac{P_t}{1-P_t}\right) = \mathbf{X}_t\boldsymbol{\theta} + s_1(x_{1t}) + s_2(x_{2t}) + s_3(x_{3t}) + \ldots + \varepsilon_t \qquad (0)$$

where $Log\left(\frac{P_t}{1-P_t}\right)$ is the logit link function for the proportion outcome as before, $\mathbf{X}_t\boldsymbol{\theta}$ is the design matrix and corresponding parameter vector (that is, any of the regression components that one wishes to continue to treat in the usual parametric fashion), $s_1(x_{1t)}, \ldots, s_p(x_{pt})$ are smoothing functions for each predictor (x) that one wishes to smooth nonparametrically, and ε_t is an error term with a binomial distribution.

We examined four GAM models on each case of the Lambert et al. (2006) data, all modeled in R using the mgcv package (Wood, 2010). Each model includes: (a) an intercept, or the participant's initial outcome level, (b) a continuous time (trend) variable X_t measured as calendar time across sessions (e.g. two sessions conducted one day apart would be 1, 2; two sessions one week apart would be 1, 8), (c) a dummy-coded treatment variable z_t (0 for baseline, 1 for treatment), and (d) a time-by-treatment interaction:

$$[X_t - (n_1 + 1)]z_t,$$

where n_1 is the time of the last data point in the first baseline phase (the data point directly preceding the initial introduction of treatment). This interaction captures the change in slope beginning at the start of treatment (Huitema & McKean, 2000).

The first GAM model is a linear model with no smoothers (Model 1):

$$Log\left(\frac{P_t}{1-P_t}\right) = \beta_0 + \beta_1 X_t + \beta_2 z_t + \beta_3[X_t - (n_1 + 1)]z_t + \varepsilon_t. \quad (0)$$

This model yields identical results to a GLM with binomial errors; it will differ somewhat from a GLM that assumes normality, the latter being a common but incorrect analysis for this kind of outcome. The second GAM model applies a smoother to the interaction term (Model 2):

$$Log\left(\frac{P_t}{1-P_t}\right) = \beta_0 + \beta_1 X_t + \beta_2 z_t + s_3([X_t - (n_1 + 1)]z_t) + \varepsilon_t. \quad (0)$$

The third model applies the smoother to the trend term (Model 3):

$$Log\left(\frac{P_t}{1-P_t}\right) = \beta_0 + s_1(X_t) + \beta_2 z_t + \beta_3[X_t - (n_1 + 1)]z_t + \varepsilon_t. \quad (0)$$

The fourth model applies a smoother to both the interaction term and the trend term (Model 4):

$$Log\left(\frac{P_t}{1-P_t}\right) = \beta_0 + s_1(X_t) + \beta_2 z_t + s_3([X_t - (n_1 + 1)]z_t) + \varepsilon_t. \quad (0)$$

Each GAM analysis gives standard regression output (i.e. regression coefficient, standard error, t-test of the coefficient, p-value) for each parametric term. For smoothed terms, output lists the effective degrees of freedom. The effective degrees of freedom are a rough measure of the complexity of the fitted spline model; edf = 1 corresponds to a linear model, and as edf increases towards edf = k, the spline model is approximately as complex as a polynomial of degree k - 1 (Hothorn & Everitt, 2010, Chapter 10). Smoothed terms also have a corresponding F-statistic and p-value that are conservative and approximate. To determine which model fits the data best, one compares various model fit statistics (R^2, deviance) along with examining significance tests (more details are in Sullivan and Shadish, 2012).

Usefulness / Applicability of Method:

Demonstration of the usefulness of the proposed methods using hypothetical or real data.

We will apply GAMs to a large set of SCDs drawn from a survey of SCD literature in 2008 (Shadish & Sullivan, 2011). We will report results on the following questions: (1) Does the conclusion about treatment effect from the usual generalized linear model change when GAMs are used to include higher order smoothing functions in the model? (2) When GAMs indicate the need for higher order smoothing functions, what degree of polynomial do they suggest? (3) How well will a simple model with linear trend do compared to a GAM with nonlinear trends, where the evaluative criterion is the size and significance of the treatment effect.

Research Design:
Description of the research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).
(May not be applicable for Methods submissions)

(Not applicable)

Data Collection and Analysis:
Description of the methods for collecting and analyzing data.
(May not be applicable for Methods submissions)

(Not applicable)

Findings / Results:
Description of the main findings with specific details.
(May not be applicable for Methods submissions)

We have already applied GAMs to about a score of cases (Shadish, Hedges, Pustejovsky, Rindskopf, Boyajian, & Sullivan, in press; Shadish & Sullivan, 2011). Results suggest treatment effects are not sensitive to trend (linear or nonlinear) about half the time, but for the rest of the cases a failure to include higher-order smoothing functions can result in incorrect conclusions about treatment effects.

Conclusions:
Description of conclusions, recommendations, and limitations based on findings.

This research should inform SCD researchers about the need to take linear or nonlinear trend into account. With further study, it may also be that GAMs may prove to be a primary analytic method for use in SCDs.